

ABSTRACT

The present invention provides a method and system to improve speech recognition using an existing audio realization of a spoken text and a true textual representation of the spoken text. The audio realization and the true textual representation can be aligned to reveal time stamps. A speech recognition can be performed on the audio realization to provide a hypothesis textual representation for the audio realization. The aligned true textual representation can be compared with the hypothesis textual representation. Single word pairs from the true and the hypothesis textual representations can be selected where the representations are different. Similarly, single word pairs can be selected from each representation where the representations are identical. A word or pronunciation database can be updated using the selected single word pairs together with the corresponding aligned audio realization.